Drowning Detection Algorithm For Intelligent Lifebuoy

1st Dasheng Yang
Zhejiang University
Hangzhou, China
yangdasheng@zju.edu.cn

2nd Yan Cao
Ningbo SUPCON Microelectronics Co.,Ltd
Ningbo, China
caoyan@nz-ic.com

3rd Yujing Feng
Ningbo SUPCON Microelectronics Co.,Ltd
Ningbo, China
fengyujing@nz-ic.com

4th Xiao Lai
ZheJiang SUPCON Technology Co.,Ltd
Hangzhou, China
laixiao@supcon.com

5th Zaisheng Pan*
Zhejiang University
Hangzhou, China
panzs@zju.edu.cn

2021 IEEE International Conference on Unmanned Systems (ICUS) | 978-1-6654-3885-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICUS52573.2021.9641291

***Abstract*—Intelligent lifebuoy is a new type of unmanned high-speed surface rescue robot, but at present it mainly relies on remote control. An intelligent lifebuoy with the ability of drownning detection, distance detection and autonomous navigation will have a good application prospect in the field of surface rescue. This paper analyzes the characteristics of drowning person and the requirement of rescue control and proposes a drowning detection algorithm for intelligent lifebuoy. An improved YOLOv4 network is designed to detect the drowning person and a geometric distance measurement method based on the bounding box is designed to detect the position. The experimental results show that the algorithm can not only achieve a high detection accuracy, but also get the distance and direction synchronously, so as to help the intelligent lifebuoy to complete a fast and effective rescue.**

Keywords—intelligent lifebuoy, surface rescue, drownning, object detection, distance measurement, neural network

I. Introduction

Personal safety in open waters is always an important issue. Drowning accidents often occur in lakes, rivers, reservoirs and coasts, etc. The traditional rescue methods include small life-saving devices such as lifebuoys and large life-saving equipment such as the rush boat or kayak. The ships and shore are usually equipped with lifebuoys and life jackets. However, all the above rescue methods and facilities must be operated by rescuers. The rescue is slow and the effective rescue time is delayed. Furthermore, there are no rescuers near the drowning people on many occasions, or even if someone find the accident, they may not be able to provide rescue due to environmental, capacity, facilities and other reasons. Drowning accidents without timely rescue seriously threaten people's lives and property.

To solve this problem, a water rescue robot has been developed, which is called Intelligent Lifebuoy, as shown in Fig. 1. Intelligent lifebuoy is a high speed water surface robot which can be operated by remote control, and specially used for rescue in ocean, river, reservoir, and so on.

Intelligent lifebuoy is generally composed of battery, driving motor, propeller and other accessories. The speed can be up to 5m/s because of the high-power brushless motor, and the large capacity lithium batteries make sure that it has a long-term endurance. So that it can quickly reach the drowning people by remote control and finish the rescue job in time.

Under the waterproof level of IP68, the intelligent lifebuoy also integrates a variety of intelligent modules, such as wireless communication, GPS positioning, loudspeaker, camera, infrared sensor and so on. So that it can realize high precision remote control, multi gear constant speed navigation, voice intercom, video acquisition and other functions. When a rescuer observes a person falling into the water, he can send a continuous movement command to the lifebuoy through the remote control, which make it move quickly to the drowning person. After the drowning person grabs the lifebuoy, he will be then quickly taken back to the shore or rescue ship. This method effectively improves the efficiency of rescue and reduces the risk, and can be greatly significant for water surface search and rescue tasks.

Fig. 1. A cruising intelligent lifebuoy.

However, there are still some shortages in the current remote control intelligent lifebuoy. First of all, at least one person needs to be aware of the drowning accident in time and ready to rescue, and the drowning person must be within the sight of the rescuer. Secondly, the rescuer must have the ability to acquire and operate the intelligent lifebuoy. These conditions are usually not satisfied in the field water environment such as lakes, rivers and reservoirs.

The primary function of lifebuoy is to rescue drowning people. Therefore, the truly intelligent lifebuoy should be able to identify the drowning person independently, and then take

978-0-7381-4657-7/21/$31.00 ©2021 IEEE

the drowning person as the object to navigate automatically as fast as possible.

With the development of artificial intelligence technology, computer vision (CV) and machine learning technology have been gradually applied to the field of perception and measurement [1] [2] [3]. The path-finding and obstacle-avoiding technologies of unmanned systems such as UAV and unmanned sailboat has also made great progress [4] [5], which make it possible to realize intelligent navigation functions on intelligent lifebuoy such as autonomous cruise, object recognition, object tracking, obstacle avoidance, one click return, etc.

In this paper, a set of drowning detection algorithm for intelligent lifebuoy is proposed. Through the images obtained by the camera, the algorithm can quickly identify the drowning accident, judge the direction and predict the distance of the accident. These data are sent to the controller to assist the intelligent lifebuoy to complete the rescue mission.

II. Related Works

A series of researches have been carried out on the detection of drowning people.

Zhang et al. used background subtraction algorithm to detect swimmers in swimming pool and discriminate the drowning and normal swimmers by comparing the pixel range of underwater swimmers [6]. Hou et al. proposed a swimming detection and tracking method based on discrete cosine transform algorithm, target detection algorithm, and traditional cam shift algorithm [7].

Peng et al. designed a drowning detection system based on the improved Mask R-CNN algorithm [8], which can judge the swimmers' postures, identify the drowning person and trigger the alarm. Jing et al. [9] also improved the Mask R-CNN algorithm by introducing the spatial attention module in the Mask branch, and designed the deep attention segmentation model SAG-Mask R-CNN, which improves the speed and accuracy of drowning detection effectively.

Alshbatat et al. developed a automated vision-based surveillance system to detect drowning incidents in swimming pools [10]. The above researches are all for the detection of drowning in swimming pool, and the systems have to be installed in a fixed position, so they are only suitable for specific environment, but not for open waters.

According to the characteristics of open waters, Chen et al. focused on the interference of the reflection of human body on the water surface to the detection of the drowning person [11]. They research the feature of watermarks, and uses HOG [12] + SVM method to recognize and detect the real drowning object. However, this method only classifies the human body and the reflection through background removal, and has no recognize ability from other interference images, and it can't evaluate of direction and distance.

Yaswanthkumar et al. [13] implemented Faster R-CNN [14] and YOLOv3 [14] algorithm on an underwater robot to identify and find the drowning person in the water, and obtained good results.

III. Design of Drowning Detection Algorithm

A. Analysis of drowning characteristics

There are typical characteristics for the scene of accidental drowning in open waters.

Firstly, unlike the swimming pools, there should be no people in open waters by default. So the intelligent lifebuoy does not need to compare the different posture of human body in the water to determine whether it is a normal swimming behavior or drowning. Any human feature detected should be judged as a person drowning.

Secondly, there are no obvious characteristics in body shape, clothing, posture or movement of the people who fall into the open water. The parts of the body exposed above the water during the struggle are also different. However, head, arm, hands and other upper body parts are more likely to be photographed.

Thirdly, people who fall into the water by accident usually struggle violently. For this kind of violent action, if the background changes little, we can usually use a variety of methods to extract the region of interest (ROI) and detect the moving object, such as three frame difference method [16], or ROI extraction algorithm based on improved YOLO [17] . However, because the lifebuoy itself is always in motion, there is no stable background. Moreover, the wild complex open waters usually have violent changes such as torrents and waves. Therefore, the traditional ROI extraction method will be difficult to apply for the intelligent lifebuoy.

Last of all, the lifebuoy needs to arrive at the place of the accident with the fastest speed and the shortest path. When approaching the drowning person, the lifebuoy should be able to slow down and park beside the drowning person. Therefore, in addition to the detection of person in the water, the lifebuoy also needs to be able to estimate the position and distance in real time.

B. Design of detection algorithm

Based on the above requirements and analysis, this paper selects the face, side face, arm, hand and other images that are likely to be collected as the key features to design the detection algorithm. People may wear life jackets in some scenes, so the life jacket is also used as a typical feature.

The detection and control algorithm structure of intelligent lifebuoy is designed as Fig. 2.

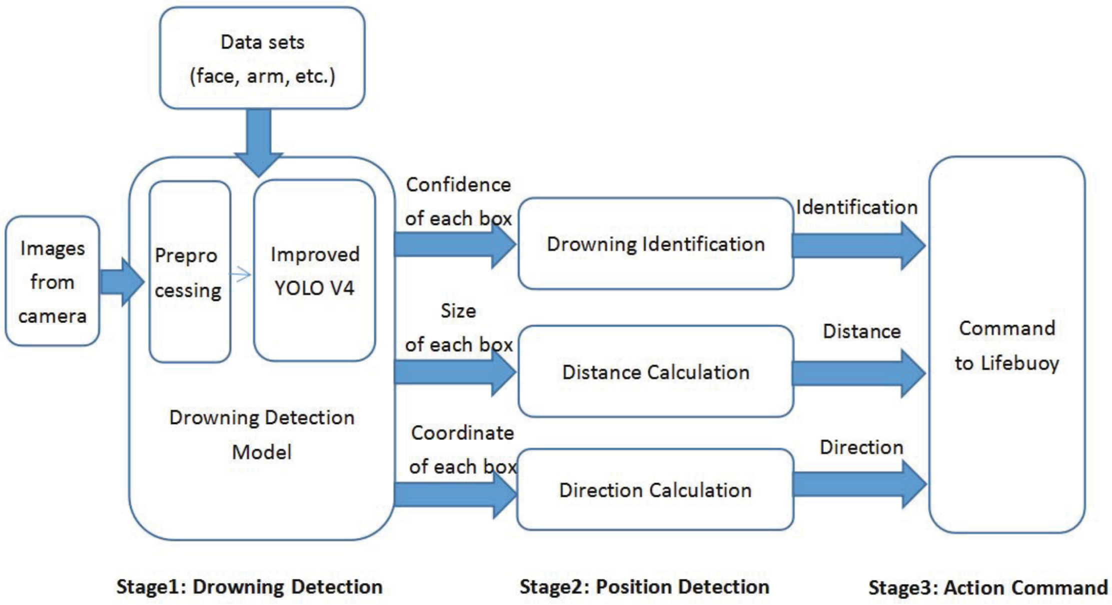

Fig. 2. The detection algorithm structure of intelligent lifebuoy.

As shown in Fig. 2, the algorithm is divided into three stages.

The first stage is the drowning detection stage. Convolution neural network (CNN) is used and a drowning detection model is obtained through a large number of training samples. The model is then deployed in the intelligent lifebuoy. The camera equipped on the lifebuoy collects image data in real time and sends it to the model for detection. The detection model identifies the head, arm, palm, life jacket or other elements in the image, and gives the confidence and coordinates of each bounding box.

In the second stage, a comprehensive confirmation of a drowning person is finished, together with the calculation of the distance and direction of the drowning person relative to the lifebuoy.

After a drowning person is confirmed, the location data obtained in stage 2 such as distance and direction are sent to the controller, and the controller switches the command to do the rescue in the third stage.

IV. Drowning Detection Stage

A. Neural network model based on improved YOLOv4

In the field of object detection, objects whose aspect ratio is less than 10% of the original image or whose absolute size is less than 32*32 are defined as small objects. Intelligent lifebuoy not only needs to be able to recognize large objects such as face and body nearby, but also needs to be able to recognize small objects such as hands more than ten meters away. Therefore, a fast detection method compatible with large and small objects is a challenge for lifebuoy.

In this paper, we choose YOLO as the basic algorithm. YOLO, which means You Only Look Once, is a real-time object detection algorithm designed by Redmon in 2016 [18]. The latest version of YOLO (YOLOv4) [19] is composed of CSPDarknet53 [20] basic network, PAN [21] feature fusion network and YOLO detection network. CSP structure is used to solve the redundant gradient information problems and reduce the amount of calculation.

YOLOv4 uses three scale feature maps for object detection, which has a certain ability of small object recognition. The low level feature map mainly reflects the details of detection object, which means an additional output scale derived from the low-level feature map layer of the YOLO backbone network can improve the detection efficiency of small objects [22].

Based on this, we propose an improved YOLOv4 network in this paper, which introduces an new output scale from the low-level feature map layer on the structure of YOLOv4, to improve the detection ability of small objects.

The structure of the improved YOLOv4 network is shown in Fig. 3.

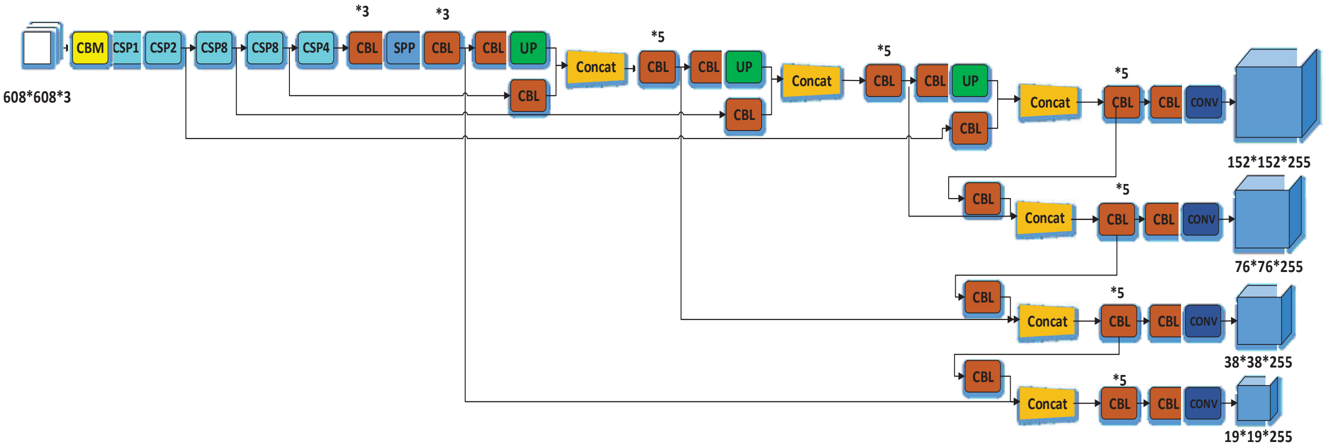

Fig. 3. The structure of the improved YOLOv4 network.

In this paper, we also improves the Non-Maximum Suppression (NMS) of YOLOv4. The function of NMS is to sort the bounding boxes of a certain type according to the confidence, set the box with the maximum confidence as the reference box, and calculate the IOU with other bounding boxes. Delete the boxes one by one if the IOU is higher than the threshold, until all redundant bounding boxes are eliminated. DIOU-NMS is used in YOLOv4, which take the overlapping area and the distance of bounding boxes into consideration.

In the application scenario of intelligent lifebuoy, the position detection is as important as person detection. The bounding box is used to do the position detection in this paper, so its accurate shoud be as accurate as possible.The DIOU-NMS algorithm used in YOLOv4 just delete of remain the bounding box depending on confidence, which is not enough to reflect the accuracy of the box itself.

Soft NMS proposed in the article of Navaneeth et al. [23] does not set the score to 0 directly when IOU is greater than the threshold, but reduces the score of the bounding box through attenuation, thus reducing the possibility of missed detection. Therefore, referring to the ideas of soft NMS and softer NMS [24], we appropriately improves DIOU-NMS of YOLOv4 in this paper. By calculating the total average of the four vertex coordinates of the bounding box whose confidence is greater than the threshold, a more accurate position information will be obtained.

B. *Data set construction*

In order to adapt to the different characteristics of drowning person, the data set is constructed based on the images of human body in water. There are three main sources of the data set. One is the portrait images in each water environment obtained from the network. The other is the images obtained by using the lifebuoy camera to take video pictures of different human postures in different environments. The third is the expanded samples by rotating or changing contrast based on previous two. The final data set is about 3000 images.

The training model adopts Pascal VOC dataset format. The open source annotation software labeling from MIT is used to annotates the image. The data set is then divided into training set, verification set and test set according to the ratio of 8:1:1.

According to the typical characteristics of the person in the water, the labels include: face, face (side), arm, hand, back and lifejacket. Because of the great difference between the front and side features of the face, they are labeled separately. In addition, life jacket is a common element in water rescue scene, so it is also used as a label.

Some examples of the dataset used in this paper are shown as Fig.

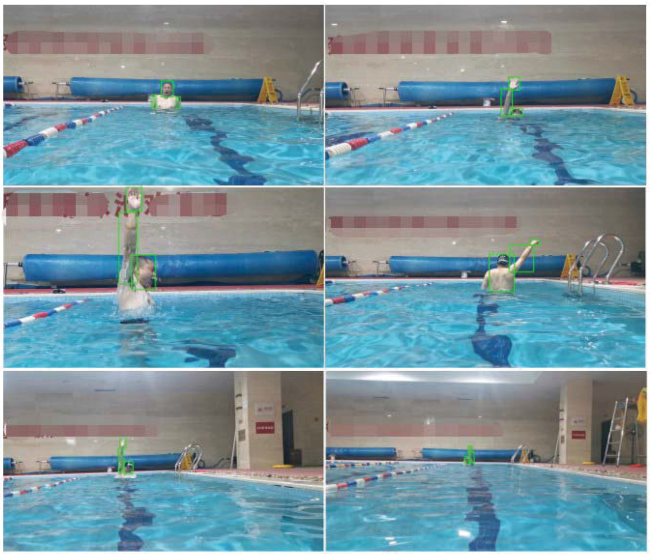

Fig. 4. Examples of the dataset.

C. *Experiment of Drowning Detection*

The experiment of training and verification is completed in a single PC environment. The training environment configuration is as follows.

- Processor: Intel(R) Core i7-6700 CPU@3.40GHz.
- GPU: NVIDIA GeForce GTX 1080.
- Operating system: Windows10.
- Software platform: Anaconda Python 3.6.
- Deep learning framework: PyTorch 0.4 + CSPDarknet53.
- Learning rate: 0.00261.

- Batch size：32.

The evaluation indexes in this test include *P* (Precision), *R* (Recall) and F_1 (F-Measure).

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{1}$$

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

$$F_1 = \frac{2*P*R}{P+R} \tag{3}$$

TP/TN/FP/FN in these equations are the numbers of samples with diffrent result after experiment. TP means True Positive, TN means True Negtive, FP means False Positive and TN means True Negtive.

In order to verify the effect of the improved YOLOv4 compared with the original YOLOv4. Two models are both trained for 40000 iterations on the same data set, and F1 values are measured on the validation set every 5000 times. The curves generated according to the experiment are shown in the Fig. 5.

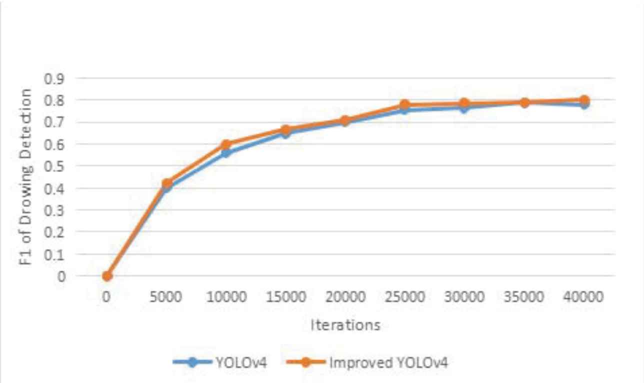

Fig. 5. F-Measure result of the improved YOLOv4.

As shown in Fig. 5 , the F-Measure is up to 0.8 after 30000 iterations. In other words, we can assume that network has more than 80% possibility to identify the human parts exposed above the water surface. Because there may be several parts that can be identified in one picture, a higher comprehensive detection rate can be obtained.The figure also shows that the improved YOLOv4 network has better effect than the original network.

V. POSITION DETECTION STAGE

A. Drowning Identification

Usually, the intelligent lifebuoy is set out with a mission. After arriving at the object water area through GPS or other information, its visual task is to find the person in the water, rather than to identify whether a person is in the water or not.

Therefore, the detection accuracy of a single body part is allowed to be lower. However, as long as multiple recognition results show that there are human characteristics in the water, it can be judged as someone falling into the water.

In this paper, the sum of all the confidence ($P_1,P_2,\ldots,P_n$) of the identified effective objects ($T_1,T_2,\ldots,T_n$) is used as the basis for drowning identification. When the total confidence is higher than the set threshold P_{thresh}, the IR (Identification Result) is true, which means the person who fell into the water has been found, and then the position will be calculated.

$$IR = \begin{cases} 1, \sum_{i=1}^{n} P_{\mathrm{n}} > P_{thresh} \\ 0, \sum_{i=1}^{n} P_{\mathrm{n}} \le P_{thresh} \end{cases} \tag{4}$$

The identification formula is as shown in For. The threshold P_{thresh} can be set after debugging in practical application. According to the experimental environment in this paper, the value is set to 0.8.

The evaluation experiments are also done and the F-Measure of drowning identification is up to 95%.

B. Direction Calculation

After the drowning person is found, the direction of the accident relative to the lifebuoy can be obtained quickly according to the distribution of the bounding box.

The camera placed at the front side of the lifebuoy has a viewing angle range ω。 The resolution of the image from the camera is R_x*R_y. If the positive center of the acquired image $R_x/2$ is set to 0 degree, which indicates the positive front direction of the lifebuoy, then the left edge of the image points to $-\omega/2$, and the right edge points to $\omega/2$ .The coordinate of the bounding box is [x_{min}, y_{man}, x_{max}, y_{max}], then the horizontal ordinate of its center point is $(x_{min} + x_{max}) / 2$.

Then the direction angle θ can be expressed as:

$$\begin{aligned} \theta &= \left(\frac{(x_{\min} + x_{\max})}{2} - \frac{R_x}{2} \right) \times \frac{\omega}{R_x} \\ &= \frac{(x_{\min} + x_{\max})}{2R_x} - \frac{\omega}{2} \end{aligned} \tag{5}$$

The schematic diagram is as shown in Fig. 6.

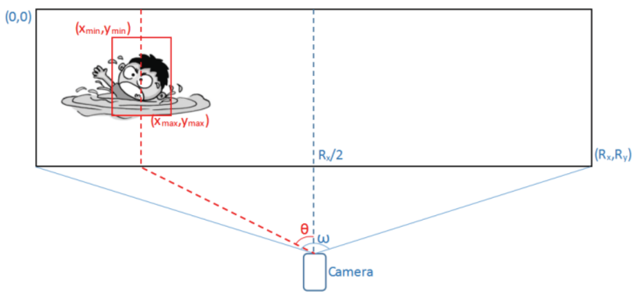

Fig. 6. The schematic diagram on direction calculation method.

Multiple objects may be detected in a image, and each object has its own bounding box to obtain the corresponding direction angle. Some of these object may have errors or deviations of bounding box. Therefore, for the detected objects, a vote will be done to eliminate the significant deviation data. The weighted mean of rest data is taken so as to get the accurate direction data.

$$\theta = \frac{\sum_{i=1}^{n} P_n \times \theta_n}{\sum_{i=1}^{n} P_n} \quad (6)$$

C. Distance Calculation

To have a effective rescue, the intelligent lifebuoy should reach the location of drowning as fast as possible. But it can not keep high speed all the time, so that the drowning person can not grasp it or even be directly knocked down and causes the second accident. So the distance between the location accident and the lifebuoy determines the speed. As discussed above, the posture of the drowning person is completely irregular, and due to the reflection of the water surface, torrent or other reasons, the laser radar or millimeter wave radar that are mature applied in auto driving may not effective on the lifebuoy anymore. Not to mention the embedded environment of intelligent lifebuoy can not bear its high integration cost.

Luckily, the main goal of intelligent lifebuoy distance calculaiton is to estimate a range of distance, which does not need to be as accurate as auto driving. In this paper, the distance and the speed are divided into several levels. When the distance is more than 10 meters, the lifebuoy moves at the highest speed; When the distance is between 10 meters and 3 meters, the lifebuoy moves at a speed of about 2 m/s; When the distance is less than 3 meters, the lifebuoy shall keep a speed of not less than 1 m/s, and move by inertia to the drowning person.

A fusion distance measurement method based on monocular vision is adopted in this paper, which can be divided into three steps. In the first step, one or more detection results are obtained by the above improved YOLOv4 network. According to the coordinates of each bounding box, a set of distance data is obtained by using geometric ranging model. The second step is to vote one or more distance data to get the distance data based on the current image. In the third step, the distance data is corrected according to the historical distance data and the speed of the lifebuoy.

The geometric ranging model based on the size of the bounding box is shown in the Fig. 7.

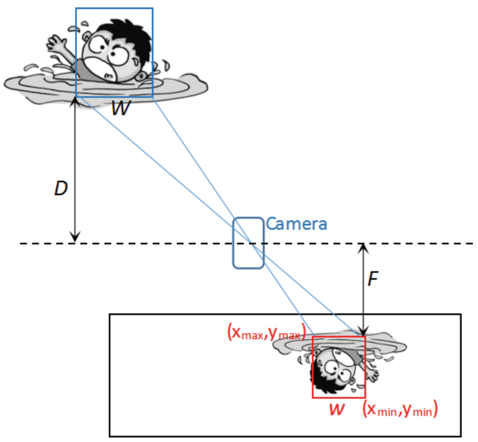

Fig. 7. The schematic diagram on distance calculation method.

As shown in Fig. 7, the coordinates of the bounding box are [x_{min}, y_{man}, x_{max}, y_{max}], so the pixel width of it is $w = x_{max} - x_{min}$. The focal length of the camera is F. The actual width of the object detected is W, and the real distance ratio between the object and the camera is D.

The imaging of the camera accords with the principle of pinhole imaging, that is：

$$D \approx \frac{W}{w} \times F \quad (7)$$

On the one hand, the bounding box of the object is not strictly linear with the object, on the other hand, it is also nonlinear due to the wide-angle characteristics of the camera.

In addition, because the bounding box of a slender object such as arm is too different at different inclinations, the real proportion cannot be expressed, in this paper, w is modified to the diagonal length of the bounding box. Correspondingly, W also take the maximum diagonal distance of length and width of object.

$$w = \sqrt{(x_{max} - x_{min})^2 + (y_{max} - y_{min})^2} \quad (8)$$

According to the above analysis, D is approximately linear with W/w and F. In order to simplify the calculation, the linear fitting is carried out in this paper.

$$D = f(w, W, F) = a \times \left(\frac{W}{w}\right) + b \quad (9)$$

In the fomula (9), a and b are the fitting coefficients obtained from the calibration fitting of the training data set. W can be obtained by querying the preset value in the controller. It can also be seen from the formula that by fitting the coefficient through calibration, the focal length F of the camera doesn't be involved in the calculation.

Similarly, when multiple objects are detected, each object has a corresponding bounding box to obtain the corresponding distance data. Therefore, voting and weighted average operation are also carried out to obtain a distance data based on current image D_{img}.

$$D_{img} = \frac{\sum_{i=1}^{n} P_n \times D_n}{\sum_{i=1}^{n} P_n} \quad (10)$$

The third step is to correct the data according to the historical distance data and the current speed and trajectory of the intelligent lifebuoy. Set the interval time between two frames as t, the speed of the lifebuoy in the previous frame as v, and the predicted distance of the previous frame as D_{last}. Because the lifebuoy is moving towards the object, the predicted distance should be

$$D_{pre} = D_{last} - v \times t \quad (11)$$

Using D_{pre} with a certain weight λ to correct the distance obtained by geometric ranging model,

$$D_{now} = \lambda \times D_{pre} + (1-\lambda) \times D_{img} \\ = D_{pre} + \lambda \times (D_{pre} - D_{img}) \tag{12}$$

According to the state equation of the intelligent lifebuoy, Kalman filtering can also be used to predict the distance data. Kalman filtering (KF) [25] is algorithm that can get the optimal estimation of the system state through the system input and output observation data. For example, Prevost et al. predicted the dynamic trajectory of UAV Based on KF [26]. The algorithm based on KF should be able to provide more accurate prediction results on lifebuoy, which will not be described in detail in this paper.

D. Experiment of Position Detection

In this paper, the camera angle range ω is 120°, and the resolution of the image is 1920 * 1080. About 100 images with real distance label are selected. After the detection through the improved YOLOv4 network described in Chapter 3, the calibration program is implemented.

The real size of the object category W is preset as the following table:

TABLE I. W TABLE

Object Type	W-Value
face	30
face(side)	30
arm	50
hand	20
back	70
lifejacket	70

The corresponding curve of D and W/w is as shown in Fig. 8.

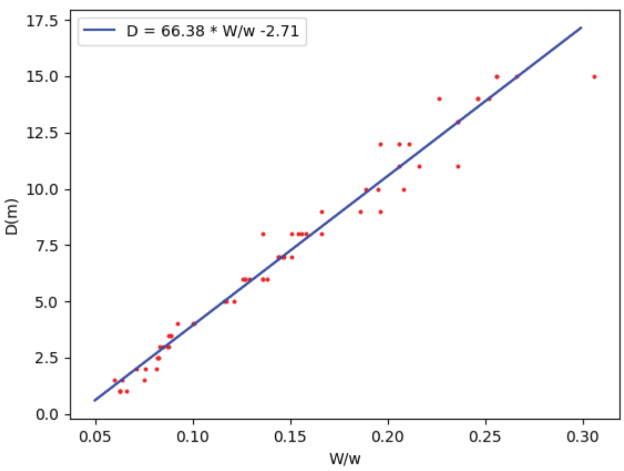

Fig. 8. The corresponding curve of D and W/w after fitting.

The calibration data are fitted and we get

$$D = 66.38 \times \left(\frac{W}{w}\right) - 2.71 \tag{13}$$

This fomula is fixed into the position detection program to realize the real-time detection of distance and direction together with the drowning detection.

Fig. 9. A example of position detection results.

As shown in Fig. 9, this model successfully identifies the hand and arm of human body. The distance of human body in water is 7.52 meters, and the direction is 39.6° to the left. Although the real distance in this image which is 6 meters has not been accurately measured, the results still fall into the range of 3~10m, and the direction angle is accurate, so that the unmanned lifebuoy can get the correct command.

We also conducted some tests in open water. Due to the strict control of the open water area, the testers can not stay in the water for a long time. However, some satisfactory results are still obtained.

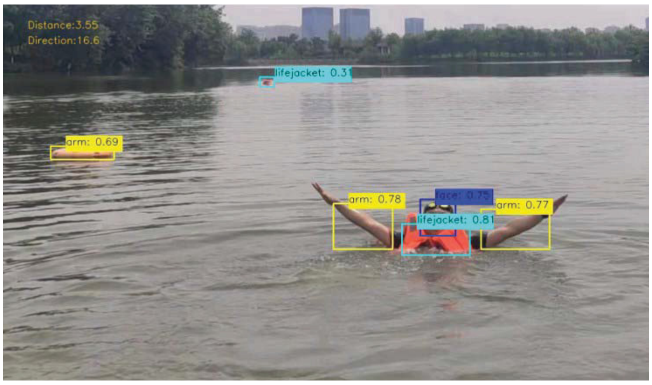

Fig. 10. A test result in open water.

As shown in Fig. 10, the algorithm accurately identified the tester in the water and estimated the distance and direction. Interestingly, the model recognized a common yellow lifebuoy not far away as an arm, which made a negative effect on the measurement results. How to improve the accuracy and prevent rescue failure due to the false identification is the focus of follow-up research of this subject.

VI. Conclusion

Aiming at the requirements of intelligent lifebuoy in open waters, and referring to some excellent methods of unmanned system, we propose a drowning detection algorithm for intelligent lifebuoy based on improved YOLOv4. Through the simultaneous detection and comprehensive evaluation of multiple typical characteristics of a person in water, a high accuracy detection of the drowning person is realized. Through the geometric distance measurement model based on the bounding box, the position of the drowning person is confirmed. The experiments show that the algorithm designed in this paper can not only identify the drowning person quickly, but also synchronously get the distance and direction of the drowning person, so as to realize the rapid rescue of the intelligent lifebuoy.

Of course, the model has still quite a lot of shortcomings.

First of all, the appearance, clothing, and the posture characteristics of person in the water are irregular, and the environmental characteristics of different open water areas are also different. It is difficult to obtain high-precision recognition ability based on the current monocular machine vision technology. The detection model trained and calibrated in this paper may not have strong generalization ability.

Secondly, the premise of the geometric ranging model based on the bounding box is that the YOLOv4 bounding box exactly frames the object, and the pose of the object matches the preset W-Value. However, the current six categories can not cover all kinds of possible situations. For example, when the arm is facing the camera, YOLOv4 may still detect the arm, but the size of its bounding box must be reduced. So the object categories need further research to improve.

Thirdly, for the low-cost embedded environment of intelligent lifebuoy, even the YOLOv4 network is too large to be supported. The compression and acceleration of the model need to be considered in the follow-up study.

In general, this paper is a preliminary try in the research topic of automatic detection and rescue of intelligent lifebuoy. On this basis, the real intelligent lifebuoy will be realized by in-depth research of the algorithm and integration with other technologies such as water surface path planning, obstacle avoidance strategy, and water bank cooperation technology, etc.

REFERENCES

[1] K. Vodrahalli and A. K. Bhowmik, “3D computer vision based on machine learning with deep neural networks: A review,” Journal of the Society for Information Display, 2017, pp.676–694.

[2] C. Patruno, M. Nitti , E. Stella and T. Orazio, “Helipad detection for accurate UAV pose estimation by means of a visual sensor,” International Journal of Advanced Robotic Systems, 2017.

[3] D. Yang, Z. Pan, Y. Cao, et al., “Wind measurement by computer vision on unmanned sailboat,” International Journal of Intelligent Robotics and Applications, 2021.

[4] D. M. Pierre, N. Zakaria and A. J. Pal, “Self-Organizing Map approach to determining compromised solutions for multi-objective UAV path planning,” International Conference on Control Automation Robotics & Vision, IEEE, 2013.

[5] H. Qiu and H. Duan, “Pigeon interaction mode switch-based UAV distributed flocking control under obstacle environments,” ISA Transactions, 2017.

[6] Y. Cui, W. Chen and F. Lei. “Swimmer detection based on background subtraction,” Journal of Shandong University, 2008, pp. 39–42.

[7] J. Hou and B. Li, “Swimming target detection and tracking technology in video image processing,” Microprocessors and Microsystems, 2021.

[8] T. Peng, J. Shen and Y. Qiao, “Design of drowning behavior detection system in swimming pool based on improved Mask R-CNN,” Transducer and Microsystem Technologies, 2021, pp. 100-103.

[9] M. Jing, T. Yu and M. Feng, “Research on drowning detection in swimming pool based on improved Mask R–CNN,” Journal of Qiaodao University (E&T), 2021, vol.36, pp. 1–7.

[10] A. Alshbatat, S. Alhameli and S. Almazrouei, “Automated Vision-based Surveillance System to Detect Drowning Incidents in Swimming Pools,” Advances in Science and Engineering Technology International Conferences, 2020.

[11] H. Chen and Q. Zhou, “Method for Judging Water Falling Personnel Based on Reflection Image Detection,” Computer Knowledge and Technolgy, 2018, vol.14, pp. 175-178.

[12] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” IEEE Computer Society Conference on Computer Vision & Pattern Recognition, 2005.

[13] M. Yaswanthku, M. Keerthana and P. Vishnu, “A Machine Vision Approach for Underwater Remote Operated Vehicle to Detect Drowning Humans,” Adv. Sci. Technol. Eng. Syst, 2020, vol. 5, pp. 1734-1740.

[14] S. Ren, K. He and R. Girshick, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, vol. 39, pp. 137-1149.

[15] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” arXiv, 2018.

[16] X. Han, G. Yuan and L. Zheng, “Research on Moving Object Detection Algorithm Based on Improved Three Frame Difference Method and Optical Flow,” Fifth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC), 2015.

[17] T. Li, K. Zhang and W. Li, “Research on ROI Algorithm of Ship Image Based on Improved YOLO,” International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), 2019.

[18] J. Redmon, S. Divvala and R. Girshick, “You Only Look Once: Unified, Real-Time Object Detection,” arXiv, 2016.

[19] A. Bochkovskiy, C. Wang and H. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv, 2020.

[20] C. Wang, H. Liao, Y. Wu and P. Chen, “CSPNet: A New Backbone that can Enhance Learning Capability of CNN,” 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020.

[21] S. Liu, L. Qi and H. Qin, “Path Aggregation Network for Instance Segmentation,” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[22] W. Jiang, B. Li and J. Gu, “A Ship Target Detection Algorithm Based on Improved YOLO V3,” Electronics Optics & Control, 2021, vol. 28, pp. 52-56.

[23] N. Bodla, B. Singh, R. Chellappa and L. S. Davis, “Soft-NMS − Improving Object Detection with One Line of Code,” 2017 IEEE International Conference on Computer Vision (ICCV), 2017.

[24] Y. He, X. Zhang, M. Savvides and K. Kitani. Softer-nms: Rethinking bounding box regression for accurate object detection,” arXiv, 2018.

[25] M. S. Grewal and Andrews A P, “Kalman filtering: theory and practice using MATLAB,” Prentice-Hall, 2001.

[26] C. G. Prevost, A. Desbiens and E. Gagnon, “Extended Kalman Filter for State Estimation and Trajectory Prediction of a Moving Object Detected by an Unmanned Aerial Vehicle,” American Control Conference, 2007.